知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum
E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于共现网络与情感分析的多平台消费者评论主题比较研究

周婷玮
上海大学管理学院 上海 200444

摘要：[**目的 / 意义**] 旨在以实证分析研究虚拟生活社区、社交平台、购物平台的用户对于同一款产品的评价内容主题倾向异同。[**方法 / 过程**] 选取“你今天真好看”App、微博、京东三个平台为实验对象，共采集 54 071 条同一护肤品的用户评论文本，采用 LDA 主题生成模型、共现网络，基于机器学习的情感分析方法对用户评论文本进行多平台比较分析。[**结果 / 结论**] 研究发现三个平台共八大评论主题的主题特征词、共现网络、主题情感上各有异同，且内容倾向符合各平台特点。

关键词：多平台比较　文本主题聚类　共现网络分析　情感分析
分类号：F724.6

引用格式：周婷玮 . 基于共现网络与情感分析的多平台消费者评论主题比较研究 [J/OL]. 知识管理论坛 , 2023, 8(2): 79-91[引用日期]. http://www.kmf.ac.cn/p/334/.

大数据移动互联网时代，用户可以在线上交流信息 [1]，以评论作为呈现形式的即时信息也被长久保留在平台上。在此背景下，用户通过共同的兴趣、职业、目标、需求等聚集在一起，形成网络虚拟社区 [2]。与实体传统社区相比，网络虚拟社区摆脱了时间和空间的限制，因而受众群体十分广泛 [1]。新媒体时代，虚拟社区营销逐渐成为电子商务领域信息传播和价值传递的主要方式。虚拟社区可以提供大量信息供消费者浏览，以满足其辅助决策的功能需求。其中，在线评论的文本内容对于消费者行为的影响起到了关键作用，但每条评论的重要性并不相同 [3]。有相关文献基于用户需求将虚拟社区的用户评论主题划分为平台推广、奖励制度、基础设施维护、专业特色、发帖规范、专业资源建设、管理团队建设、社会对接和用户互动等 [4]。但该类研究主要基于虚拟社区建设，研究用户参与的特点与影响，对于虚拟社区内的用户评论研究则聚焦单一虚拟社区，缺乏关于虚拟社区特色化的用户评论内容主题倾向以及跨平台比较研究。

本研究采集了“你今天真好看”App、微博、京东三种渠道中同一护肤品类商品的评论数据共计 54 071 条，将护肤虚拟社区、社交平台、购物平台以护肤专题为例进行对比，并采用 LDA(Latent Dirichlet Allocation) 主题生成模

作者简介：周婷玮，硕士研究生，E-mail：zhoutingwei99@126.com。
收稿日期：2023-01-28　　发表日期：2023-03-24　　本文责任编辑：刘远颖

型、共现网络 (Co-Occurrence Network) 和情感分析对用户评论文本进行多平台比较研究。本文旨在分析各平台评论主题倾向异同，比对同一主题下的内容差异，以期为品牌方企业在新品配方设计、已有商品优化、用户适用人群定位、用户产品需求获取、实际使用效果反馈、产品代言人选择、产品定价等方面提供精准的获取渠道决策参考，并为相关商业应用提供参考。

1 相关研究

1.1 虚拟社区与用户

目前，对于虚拟社区的相关研究主要集中于虚拟学术社区建设和虚拟品牌社区中的用户参与。周阳等[5]利用质性分析软件 NVIVO 12.0 对小木虫论坛回帖内容进行编码，据此构建虚拟学术社区参与行为的模型，发现不同等级的用户的参与行为具有不同的特点。C. Huaruo 等[6]发现虚拟学术社区持续知识共享意愿与知识共享满意度呈显著正相关，而后者受启发式因素（知识共享数量、知识源可信度）和系统因素（知识共享质量、知识共享有用性）的影响。M. Tabish 等[7]使用偏最小二乘法 (Partial Least Squares，PLS) 进行分析，研究认为虚拟社区参与对品牌信任有积极作用，而品牌信任将会对品牌选择产生积极影响。

为了深入挖掘虚拟社区用户评论中隐藏的用户需求，相关文献对虚拟社区的用户评论文本进行分析。虽然在线评论的文本内容对于消费者行为的影响起到了关键作用，但根据陈秀秀[3]的研究，每条评论的重要性并不相同。更有研究发现了用户需求与品牌之间的联系，并挖掘出含用户需求内容的评论文本的价值。例如，P. Cara[8]在研究消费者对在线评论来源可信度和有用性的感知如何影响与零售商店网络社区接触的意向和购买意向时发现，消费者希望在购物情境中与零售店网络社区进行接触。U. Chakraborty[9]利用结构方程模型的 bootstrapping 方法来检验品牌资产维度在购买意愿与在线评论之间的中介效应，最终发现营销人员应该更多地关注品牌知名度和感知价值，因为这两者最终影响消费者的购买意愿。屈慧君[10]针对微博品牌社区中用户参与动机、互动活力、信任、购买意愿之间的两两关系进行研究，发现这四者互成正相关关系。

综上所述，当前围绕虚拟社区相关的研究文献内容主要包含两个方面：专业虚拟社区中不同用户参与行为的特点、参与意愿和对虚拟社区的影响；对用户评论内容进行文本挖掘分析，研究其中潜藏的用户对虚拟社区和品牌的需求，以及用户信任、购买决策等的影响因素。然而对于用户参与在商业方面的实际应用价值定义较为笼统，且用户对于品牌产品的使用反馈是其共创价值的重要体现，而对此中用户参与进行深入研究的文献较少。对于虚拟社区评论的研究大多针对虚拟品牌社区，对虚拟生活社区的研究少有涉及。

1.2 主题挖掘研究：基于 LDA 的主题共现网络与情感分析

现有文献中对于文本中主题聚类通常采用两种方式，一种是根据文本相似度的计算进行聚类的传统模型[11-12]，其作用于原始文本，以两个文本的字符匹配程度或距离作为相似度的衡量标准，其优点是算法原理简单且易于实现，但依赖词向量空间距离[13]，存在文本向量维数过高以及语义敏感度差的问题[14]；另一种则是使用 LDA 主题生成模型。

LDA 模型最早由 D. M. Blei 等[15]提出，是一种基于贝叶斯概率的、非监督机器学习技术，属于自然语言处理的两大推力模型之一，可以用来识别大规模语料库（corpus）中潜藏的主题信息。在 LDA 主题生成模型中，每份文本被表示为一系列主题的多项分布，而每个主题则被表示为一系列单词的多项分布[16]。LDA 主题生成模型由于具有较好的处理大规模语料的能力、降维能力，成为了近年来主题挖掘研究中的一个热门方向[17]。

该模型可应用于多种情境。在历史学领域，何琳等[18]对《左传》等史书中的事件触发词借

知识管理论坛，2023(2): 79-91
DOI: 10.13266/j.issn.2095-5472.2023.007

助 LDA 实现了其自动识别和分类；在司法实践领域，王慧等 [19] 使用 LDA 提取案情三元组信息，解决了现有司法数据分析方法与案件数据的内部信息关联不够、低效落后的问题；在社交平台舆情管控领域，张雷等 [20] 基于 LDA 模型构建高校师德舆情下微博用户主题生成模型，通过识别主题最优传播路径进行舆论引导，对突发的舆情实现预测和管制优化；在电商领域，崔宁等 [21] 以 LDA 主题模型和偏序集理论为基础，对大量在线商品评论构建在线商品评论分析模型，帮助商家掌握顾客消费需求和潜在倾向。综上可见，LDA 主题生成模型应用广泛，且能高效、准确地挖掘出文本的主题特征。

基于机器学习的情感分析作为有监督学习，需要人工对文本库进行正负样本标注，再选择合适的算法对分类器进行训练 [22]。自然语言处理的分类算法主要有朴素贝叶斯算法、KNN 算法、支持向量机算、神经网络算法等。研究表明用机器学习方法对 10 万条以下的文本信息进行分类时，采用朴素贝叶斯方法能达到较好的效果 [23]。本研究将采用朴素贝叶斯算法训练分类器，对三个平台的文本情感正负极性进行处理。

2 研究方法

2.1 数据来源

本研究选取 2020 年 6 月—2022 年 3 月护肤虚拟社区、社交平台、购物平台的用户评论，并选取了“芙丽芳丝净润洗面霜”作为评论数据的采集目标对象。根据 iMedia Research（艾媒咨询）数据显示 [24]，90.4% 中国网民使用洗面奶清洁面部，这表明在消费者群体中洗面奶的使用极为普遍。而 2021 年京东“618”促消活动中，芙丽芳丝净润洗面霜拿下洁面品类销量第一，且于护肤虚拟社区“你今天真好看”App 护肤品洁面榜单中，常年保持在综合排行榜第一。因此选择“芙丽芳丝净润洗面霜”作为采集目标对象得到的研究结论具有代表性和普遍性。具体数据来源如下：

（1）护肤虚拟社区的用户评论文本爬取自“你今天真好看”App。“你今天真好看”App 于 2016 年底上线，是世界上首款通过素颜照即可进行肤质检测的手机端应用。作为一个专业的护肤品虚拟社区，其在提供肤质检测的同时，根据检测结果为用户推荐合适的护肤品，并且拥有 200 多万种护肤品的产品库、1 500 多万条产品评价、1 290 万的用户群体，截至 2023 年 1 月已提供超过 2.37 亿次拍照测肤服务（数据来自 App 官网）。本研究爬取“芙丽芳丝净润洗面霜”产品下的用户评论数据，最终得到 31 359 条有效评论文本。

（2）社交平台的用户评论文本爬取自微博。微博是一种基于用户关系进行信息分享、传播以及获取，通过关注机制分享简短实时信息的广播式社交媒体与网络平台。截至 2021 年 9 月，微博月活跃用户达 5.73 亿，来自移动端比例为 94%；日活跃用户数达到 2.48亿 [25]。本研究爬取“芙丽芳丝洁面”话题下用户博文数据，最终得到 12 404 条有效评论文本。

（3）购物平台的用户评论文本爬取自京东。京东作为中国自营式电商企业，在线销售包括化妆品与其他个人护理用品在内的商品共 13 大类 3 150 万种 SKU(Stock Keeping Unit，库存量单位)。本研究爬取“芙丽芳丝净润洗面霜”产品下用户评论数据，最终得到 10 308 条有效评论文本。

本研究从以上三个平台中获得共计 54 071 条文本数据。

2.2 评论文本主题生成模型

本研究的主要流程分为以下几个步骤：

（1）在“你今天真好看”App、微博、京东三个平台分别使用 Python 爬虫获取“芙丽芳丝净润洗面霜”的评论文本，形成多平台评论文本库，为后文的文本分析做好准备工作。

（2）对评论文本进行预处理，即加入领域专有词，分词，去除停用词，标注词性。

（3）使用 LDA 主题生成模型对三种来源的评论文本分别进行聚类，根据困惑度得出最

佳主题数，并由人工对各主题进行定义。

（4）将各平台用户评论文本中聚类得出的前 150 个特征词作为关键词，用于研究各主题内及主题间的共现网络，并对评论文本与主题进行情感分析。

（5）对比三个平台的共现网络与情感分析结果，探索各平台的侧重项。

具体流程如图 1 所示：

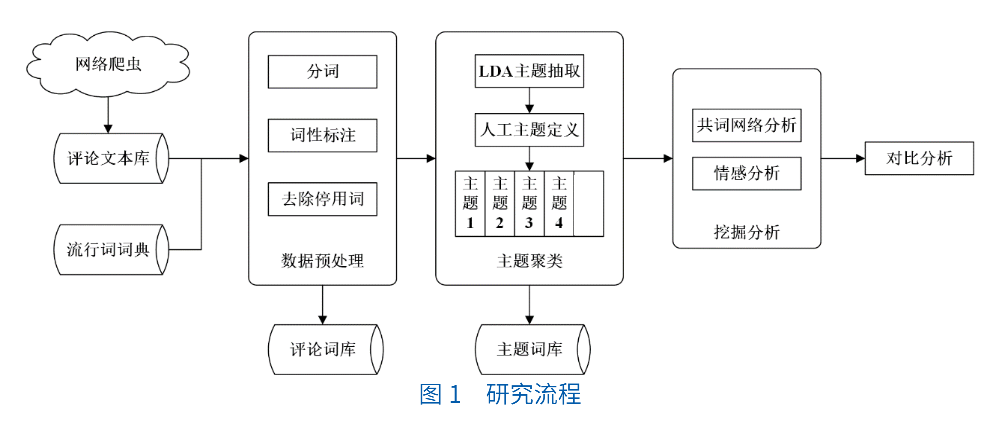
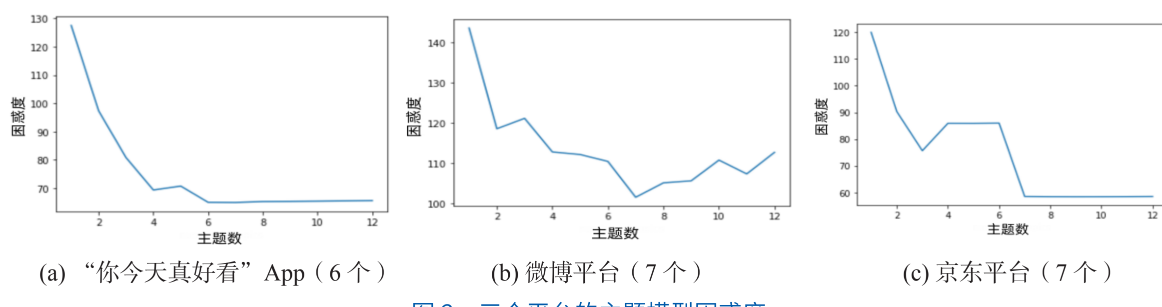


图 1　研究流程

本节将对研究流程中较为重要的部分进行详述。

（1）数据预处理。在获取评论文本后，参照哈工大停用词表 (767)，对文本进行去除停用词、标注词性、分词。为了确保本研究分词效果的完整性与准确性，加入了阅读全部评论后人工定义的流行词表 (97)，以协助补全 Python 的第三方包 Jieba 在对文本进行分词时使用的词库。

（2）LDA 主题生成模型。本研究使用 Python 的 Latent Dirichlet Allocation 函数对各平台用户评论文本分别进行 LDA 聚类并使用第三方包 pyLDAvis 对主题生成结果进行可视化分析，以避免过拟合的问题出现。依据 LDA 聚类结果并参照折肘法 [26] 寻找三个平台的困惑度拐点，得出的最佳主题数见图 2。根据所有平台的结果，单个主题下的特征词去重后，构建三者的特征词总库。根据词频分别对各平台进行特征词排序，得到每个平台各自的特征词词频表。



(a) “你今天真好看”App（6 个）　(b) 微博平台（7 个）　(c) 京东平台（7 个）

图 2　三个平台的主题模型困惑度

（3）主题共现网络分析。在 KH Coder 中经多次实验，网络图成员为 150 时作图效果最佳，因此依据 LDA 的聚类结果，在特征词词频表中取前 150 个单词进行网络图绘制。经过验证发现，三个平台有主题编码的单词词频在总词频占比分别为“你今天真好看”App 85.36%、微博 95.12%、京东 92.7%，结果具有代表性。最后，根据各主题特征词及其在同一条评论文本中的共现关系，对主题名称进行定义。使用 KH Coder 软件通过将多个主题特征词以同一主题名称进行同义词替换的方式对特征词给予主题编码，以此为基础可研究主题之间的共现关系，即不同主题

知识管理论坛，2023(2): 79-91
DOI: 10.13266/j.issn.2095-5472.2023.007

特征词在同一条评论文本中的共现现象。采用余弦距离进行文本中词语与词语，主题与主题的共现分析与语义网络构建。主题—主题的共现矩阵示例见表 1，对角线元素代表该主题在从该平台获取的用户评论文本中出现的总频次，非对角线元素代表两主题在同一条评论文本中共现的总次数。本文使用了 KH Coder 和 Gephi 软件对各平台的共现网络的可视化进行绘制工作，采用最小生成树法提取出共现网络的主干绘制共现网络，以达到精简网络图，使结果更为直观的目的。

表 1　主题—主题的共现矩阵示例

主题	1产品配方	2适合肤质人群	3产品价格	4用户使用感受	5用户购买原因	6商家活动
1	15 032	5 125	505	12 852	6 927	432
2	5 125	8 420	238	6 768	4 078	257
3	505	238	960	626	385	39
4	12 852	6 768	626	20 667	8 062	619
5	6 927	4 078	385	8 062	9 988	599
6	432	257	39	619	599	816

注：此表基于“你今天真好看”App的评论文本数据

（4）主题情感分析。根据主题聚类与定义的结果，依据文献 [27] 对词语情感极性的划分，分为正面、中性、负面并标注为 1、0、-1，以便于训练与统计，并使用 Tf-idf 方法提取特征，以达到突出重要单词、抑制次要单词的目的 [22]。训练数据方面，采用朴素贝叶斯方法，根据已知的先验概率和条件概率，基于词—类组合概率计算后验概率的所属类别 [23]。通过转化后的词向量对分类器进行训练，训练集与测试集的比例为 8:2，并使用 Python 中的第三方包 Sklearn 评价准确率，得到“你今天真好看”的结果准确率为 0.820 1，微博为 0.781 2，京东为 0.805 9。

3 实证分析

3.1 LDA 主题聚类分析

本研究使用 Python 进行 LDA 聚类，根据主题聚类的结果，总结出三个平台下共 8 个主题，其中：产品配方（成分 / 肤感）、适合肤质人群（用户皮肤问题）、产品价格、用户使用感受（使用体验、使用搭配）、用户购买原因（为什么购买、竞品、回购意愿）、商家活动 6 个主题为三个平台的共有主题。除此之外，产品代言人为微博平台特有的主题，购买平台 / 物流为京东平台特有的主题。LDA 主题聚类结果如表 2 所示：

表 2　三个平台各主题下词频前五的特征词

序号	主题名称	“你今天真好看”App	微博	京东
1	产品配方	洗面奶 味道 清洁力 泡沫 效果	氨基酸 洗面奶 洁面 泡沫 洁面乳	洗面奶 泡沫 氨基酸 味道 清洁
2	适合肤质人群	敏感肌 油皮 干皮 痘痘 油	敏感肌 干皮 油皮 肤质 油	敏感肌 肤质 程度 油皮 干皮
3	产品价格	价格 学生党 缺点 贵 毛病	到手 价 大洋 价格 史低	价格 优惠 活动价 到手 价位
4	用户使用感受	味道 脸 清洁力 皮肤 感觉	肌肤 清洁 面膜 脸 面霜	包装 效果 皮肤 产品 脸
5	用户购买原因	洗面奶 效果 氨基酸 假滑 使用感	芙丽芳丝 氨基酸 洗面奶 洁面 品牌	洗面奶 芙丽芳丝 效果 产品 氨基酸
6	商家活动	时候 水乳 脸部 双十一 小样	链接 网页 洁面乳 水乳 到手 活动	活动 小样 双十一 时候 礼盒
7	产品代言人	无	杨超越 代言人 海底 英雄 天真蓝	无
8	购买平台/物流	无	无	物流 东西 速度 正品 快递

部分特征词出现在了两个主题下，但考虑到主题定义，特征词在各主题下实际含义并不重叠，因此保留。对三个平台下各主题特征词在该平台评论文本库中的词频总占比进行统计分析（即经主题编码处理），如图 3 所示，得出各主题在该平台评论中的占比。由于存在主题共现现象，因而各平台下各主题占比之和超出百分之一百。选取在各平台中均占有较高比重的四项主题，通过词云对比用户评论文本实际内容进行分析。

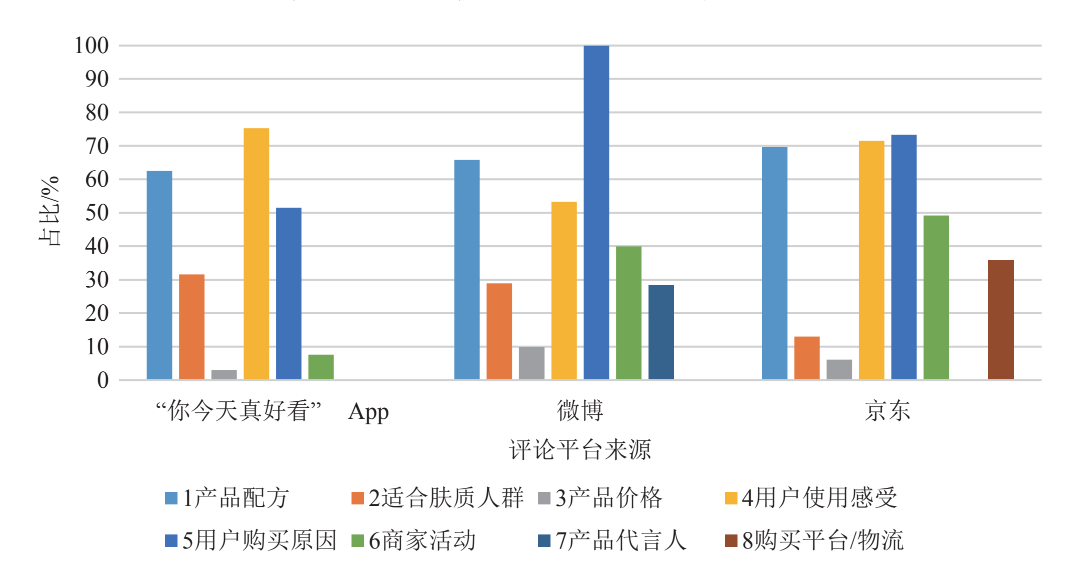
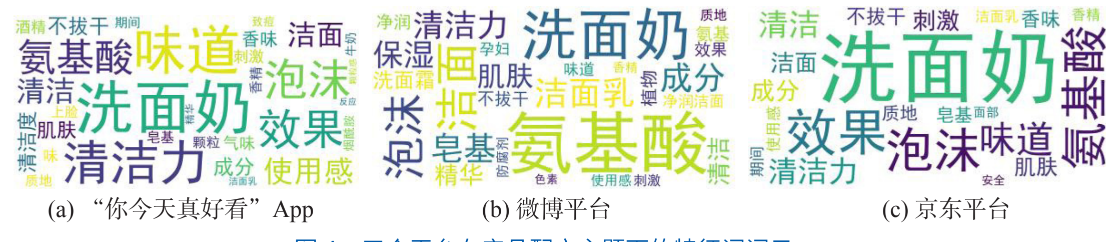


图 3　各主题在该平台评论中的占比

（1）产品配方（成分 / 肤感）主题下的特征词词云见图 4。三个平台的用户评论对于洗面奶本身的基本配方属性如氨基酸与皂基，该类配方对应的洗面奶的基本功能属性（即清洁度）以及该款产品作为氨基酸类洗面奶的特性（即洗完皮肤不拔干的效果）都有较高的关注。而“你今天真好看”App 在洗面奶基本属性以外的其他功效成分更为关注；同时对于成分是否对皮肤具有刺激性、致痘可能性的聚焦是另两个平台所没有的。微博平台的用户则更关注孕妇是否能够使用该款产品，包括成分是否提取自植物，是否含有香精、色素或防腐剂等。前两个平台与京东平台在该主题下特征词的被包含关系。



(a)“你今天真好看”App　(b) 微博平台　(c) 京东平台

图 4　三个平台在产品配方主题下的特征词词云

（2）适合肤质人群（用户皮肤问题）主题下的特征词词云见图 5。“你今天真好看”App 的特征词明显多于并包含了另两个平台在该主题下的特征词。该 App 对于脸部的区域划分更为细致，如脸颊、额头、鼻子、t 区；对于肤质和皮肤已有问题的描述更为具体，不再仅以毛孔、油、敏感肌等概括，增加了黑头、痘痘、混油（皮肤部分区域为中性皮肤，部分为油性皮肤）、油痘肌等；比之另外两个平台，增加了有关适用气候环境的描述，如换季、秋冬、时期等。

知识管理论坛，2023(2): 79-91
DOI: 10.13266/j.issn.2095-5472.2023.007



(a) “你今天真好看”App　　(b) 微博平台　　(c) 京东平台

图 5　三个平台在适合肤质人群主题下的特征词词云

（3）用户使用感受（使用体验、使用搭配）主题下的特征词词云见图 6。三个平台的用户评论均对于洗面奶基本功效（即起泡能力和清洁能力）给予了较大的关注。“你今天真好看”App 的特征词明显多于另两个平台在该主题下的特征词。“你今天真好看”App 倾向使用体验，对于产品的各种功效覆盖较全，包括控油、水油平衡、减轻皮肤负担、颗粒物添加以及使用后皮肤出现的问题（如爆痘、起闭口、对眼睛较为刺激等）；微博平台倾向使用搭配，包括与起泡网一起使用，以及后续搭配雅诗兰黛、cpb 等品牌的面膜、面霜等其他类目护肤品；京东平台则对于产品外包装的盒子、产品容量、质量更为重视。





(a) “你今天真好看”App　　(b) 微博平台　　(c) 京东平台

图 6　三个平台在用户使用感受主题下的特征词词云

（4）用户购买原因（为什么购买、竞品、回购意愿）主题下的特征词词云见图 7。三个平台用户购买该产品的主要原因均集中于该产品的主要特征属性，如品牌——芙丽芳丝、护肤品类目——洗面奶、主打成分——氨基酸、主要功能——洁面等，并且均提到了该产品的竞品品牌，如旁氏、珂润、elta 等。“你今天真好看”App 对于回购意愿的表达较为积极，初次购买的原因多见于受网络媒体传播的影响；微博与京东平台初次购买或使用原因则是来自周围亲友的推荐或赠送、同平台同商品下的评论以及对品牌的信赖。





(a) “你今天真好看”App　　(b) 微博平台　　(c) 京东平台

图 7　三个平台在用户购买原因主题下的特征词词云

3.2 关键词与主题共现网络分析

各平台的关键词共现网络见图 8，图中气泡大小代表词频，连线深浅度代表边的权重。共现的颜色分区与 LDA 主题聚类效果基本保持一致。

(a) “你今天真好看”App

(b) 微博平台

(c) 京东平台

图 8　三个平台的关键词共现网络

知识管理论坛，2023(2): 79-91
DOI: 10.13266/j.issn.2095-5472.2023.007

“你今天真好看”App 的用户评论中，高词频并与多个词有连接的关键节点相对另两个平台较多，产品配方与使用体验之间的联系最受关注。微博平台具有产品代言人这一特色主题。之所以会出现图 3 中微博平台用户购买原因中主题占比高达 99.9% 这一异常现象，是由于微博作为中国主流的公共社交媒体平台，品牌方在该平台上投放了大量该产品的广告，导致品牌与产品名称相关的单词词频异常偏高，用户购买原因主题占据了极高的比重，可见微博在营销方面的特征属性远超出其他属性。京东平台具有购买平台 / 物流这一特色主题。根据关键词之间的关系可以得出，用户在这一主题下最为关注卖家、快递的服务态度与物流的送货速度。作为一个购物平台，京东的关键词共现网络图中也出现了三个平台中仅有的价格与“双十一”活动、“送人”与“外包装”、“网上”与“评论”的强关联。

各平台的主题共现网络见图 9，仅以连线的粗细表示主题之间共现的频率。①用户使用感受是“你今天真好看”App 最为核心的主题，其与产品配方之间的共现频率最高，而产品配方与用户使用体验、用户购买原因、适合肤质人群四者之间均存在较强两两关联的关系，这再次证实了用户对于产品配方与使用体验之间的联系最为关注，并且构成了用户购买的原因，而用户使用感受也会因用户是否属于产品配方对应的适合肤质人群而受到影响。②微博平台中，产品代言人主题虽然在词频上具有较高的占比，但是整体较为孤立，仅构成用户购买意向的小部分原因。商家活动取代了适合肤质人群成为了核心主题群中的第四个元素，且其与用户购买原因高频共现关系与图 7 中该平台下产品价格主题为三个平台中占比最高的统计结果相联系，表明微博平台用户对于商家活动与价格具有更高的信息传播热情。③京东平台中的特色主题——购买平台 / 物流与其他几个主题均构成了较强的共现关系，是由京东平台本身特有的购物平台基本属性以及其在购物平台中自营物流的特征属性造成的，并非是因为其他主题会受到购买平台 / 物流这一主题的影响。

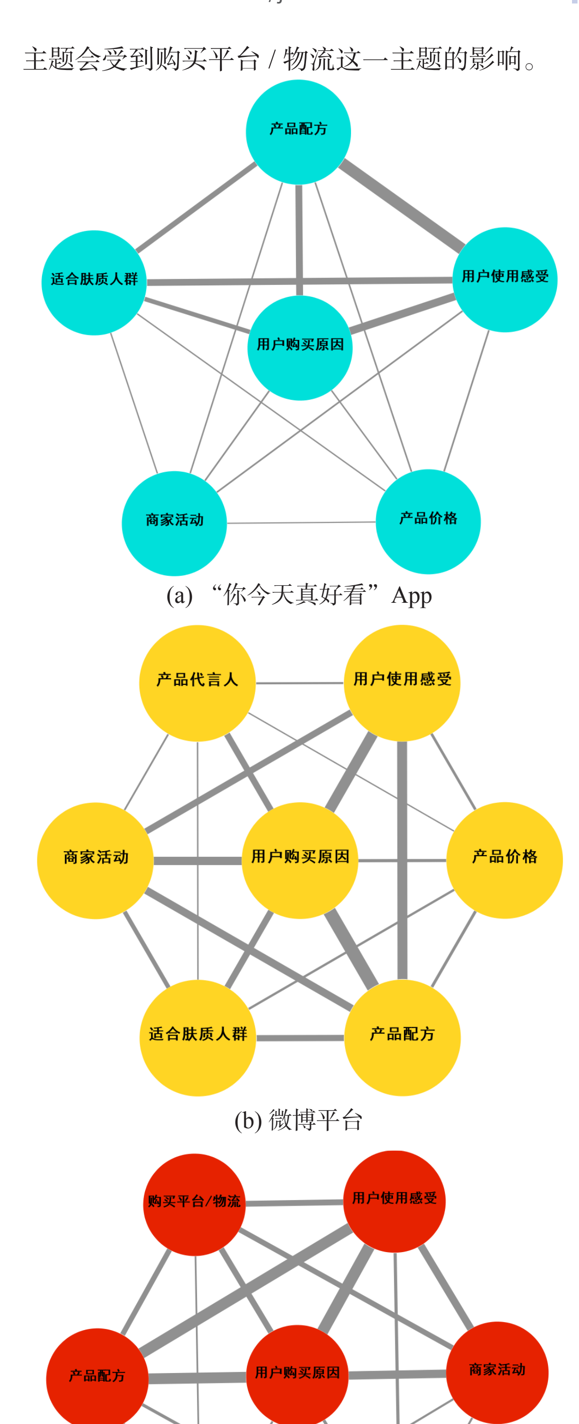


图 9　三个平台的主题共现网络

3.3 主题情感分析

本研究得出三个平台各主题的情感分布，见表3。以总体占比为权重，计算得出三个平台各主题的正面、中性、负面的加权平均情感强度，见图10。

表 3　三个平台情感分析结果

主题	“你今天真好看”App				微博				京东			
	总体占比/%	正面	中性	负面	总体占比/%	正面	中性	负面	总体占比/%	正面	中性	负面
产品配方	26.90	0.75	0.09	0.15	20.16	0.08	0.86	0.06	21.85	0.94	0.04	0.02
适合肤质人群	15.07	0.74	0.10	0.16	8.86	0.10	0.83	0.07	4.08	0.89	0.08	0.02
产品价格	1.72	0.44	0.05	0.51	3.05	0.10	0.86	0.05	1.92	0.95	0.05	0.00
用户使用感受	36.98	0.76	0.10	0.14	16.34	0.08	0.85	0.07	22.45	0.93	0.04	0.02
用户购买原因	17.87	0.75	0.10	0.15	30.62	0.06	0.89	0.05	23.01	0.93	0.05	0.02
商家活动	1.46	0.74	0.08	0.18	12.25	0.08	0.86	0.06	15.45	0.95	0.03	0.02
产品代言人	0	0	0	0	8.71	0.01	0.96	0.03	0	0	0	0
购买平台/物流	0	0	0	0	0	0	0	0	11.24	0.94	0.03	0.03
情感强度	0	0.75	0.10	0.15	0	0.07	0.79	0.05	0	0.83	0.04	0.02

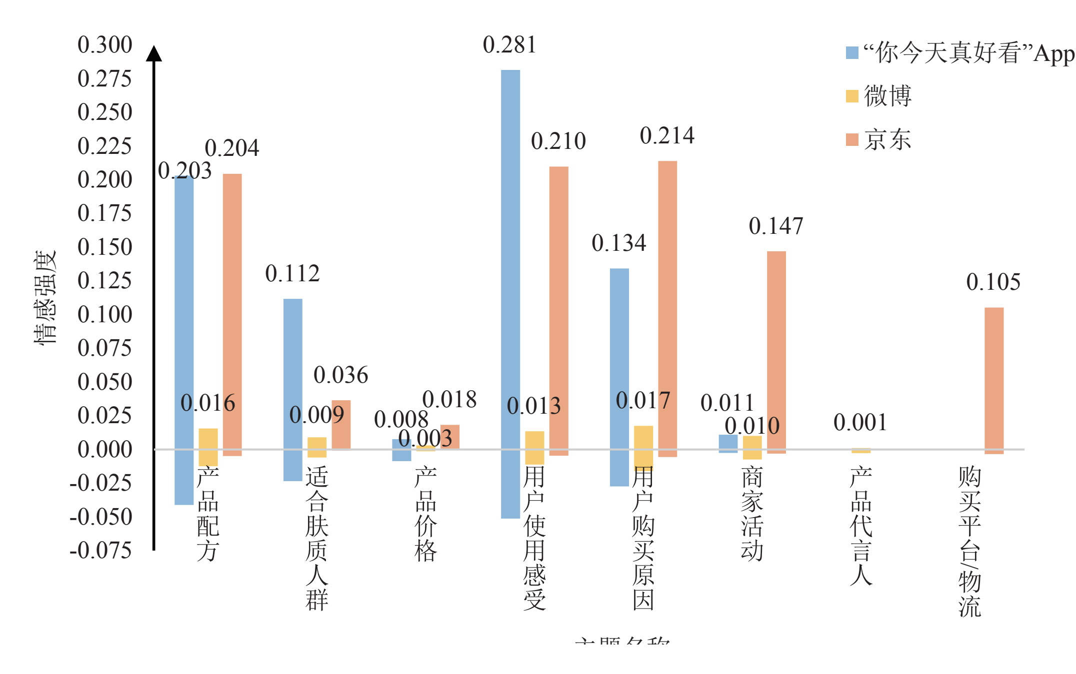


图 10　三个平台各主题的情感强度

综合8个主题的情感强度，“你今天真好看”App的三极总情感强度分别为0.75、0.1、0.15，微博平台的三极总情感强度分别为0.07、0.79、0.05，京东平台的三极总情感强度分别为0.83、0.04、0.02。微博的中性评论占比极高，是由于在机器学习过程中将模型训练数据的营销推广类文本的情感极性设为中性，这也证实了微博用户评论文本中含有大量营销投放内容。而京东的情感值正面占比最高，是由于京东作为购物平台，其用户具有“习惯性好评”的特有属性，这也是在评论文本中大量出现一组词。京东平台的中性评论则相对略高则是因为类似“敏感皮肤应该可以用”“还可以”“还行”等的评论较多，体现了该平台用户对于详细描述自身皮肤问题的积极性相对较低。“你今天真好看”App的用户评论文本所有主题的

ChinaXiv合作期刊
知识管理论坛，2023(2): 79-91
DOI: 10.13266/j.issn.2095-5472.2023.007

负面情感比例均偏高，是由于负面评论主要集中在使用产品会导致的皮肤问题。值得注意的是，产品价格在该平台中的负面评论占比高达0.51，即过半用户认为该产品价格偏高，这是因为根据百度指数统计，截至2022年4月7日，不同于微博有83.21%、京东有85.94%的用户处于20—39岁，该平台用户显示出年轻化特征，73.87%的用户年龄在29岁以下。该年龄段的用户经济水平较低，而“芙丽芳丝净润洗面霜”100g/125元的价格对于该年龄段的人群而言属于较高的消费水平。总体而言，三个平台各主题情感的波动基本均趋于稳定。

4 结论

4.1 主要发现

本研究比对了护肤虚拟社区、社交平台、购物平台的评论文本，根据LDA主题生成模型的聚类结果将用户评论文本分为产品配方（成分/肤感）、适合肤质人群（用户皮肤问题）、产品价格、用户使用感受（使用体验、使用搭配）、用户购买原因（为什么购买、竞品、回购意愿）、商家活动6个共有主题，以及产品代言人、购买平台/物流这两个社交平台和购物平台的特有主题。总结归纳以下两点主要发现：

（1）护肤虚拟社区对于产品配方、适合肤质人群、用户使用感受、用户购买原因这6个主题的评论内容描述最为详细，涵盖面最广，是品牌方获取用户需求及其购买决策的影响因素的最佳途径。社交平台中汇集了大量的广告信息，使得真实的用户反馈挖掘较为困难。购物平台用户对于在评论中表达积极情绪的意愿较高，但详细描述自身皮肤问题的积极性相对较低，即隐私感较强，分享欲较低。

（2）产品配方、用户使用体验、用户购买原因等3个主题在三个平台中均构成了较强的关联性，是三个平台核心主题组的共有组成部分。产品代言人主题虽然在词频上具有较高的占比，但是整体较为孤立，仅构成用户购买意向的小部分原因。三个平台各自的核心主题组中，护肤虚拟社区的适合肤质人群、社交平台的商家活动、购物平台的购买平台/物流与上述三个主题分别构成四元素核心主题组，从侧面体现了三个平台各自的特征：护肤虚拟社区中护肤品用户的反馈全面且针对性强，社交平台购买“水军”发布广告信息具有更广的消费者触达效果，购物平台拥有商品送达这一额外服务，尤其是拥有自建物流的京东平台，使得这一主题更具有讨论度。

本文的研究贡献包括：①将虚拟社区的用户评论文本与社交平台、购物平台同一产品下的用户评论文本通过主题聚类、关键词共现网络、主题共现网络、情感分析进行对比研究，得出了内容倾向的异同；②填补了对虚拟生活社区的研究少有涉及的空缺；③对用户参与在商业方面的实际应用价值给出了8个主题，亦可作为衡量指标构建用户评论文本价值评定模型进行用户价值共创方面更为深入的研究。

4.2 结果应用

当下，护肤品行业正经历着重大改革，自2021年1月1日起施行的《化妆品监督管理条例》中强调了产品功效的实证和配方的安全性。因此，除护肤品行业已有的牛角膜混浊和通透实验（BCOP）、鸡胚绒毛尿囊膜试验（HET-CAM/CAMVA）等产品安全性实验，品牌方企业在进行新品配方设计或已有商品优化时，应正视并积极获取护肤虚拟社区中对于用户已有的皮肤问题与反馈产品使用问题的负面情绪评论文本。考虑到用户对于产品配方与使用体验之间的联系最为关注，并且构成了用户购买的原因，而用户使用感受也会因用户是否属于产品配方对应的适合肤质人群而受到影响，品牌方企业需给予足够的重视并进行相应的配方改良，或注明适用人群，从而从根源上降低负面评论的出现频率，提高产品用户满意度与品牌信任度，从而提高产品销量。

购物平台本身与消费极为密切的特性决定了该平台的用户获取商家活动与价格变动信息十分便利，而护肤虚拟社区用户虽然对产品价

格较为敏感，却对商家活动主题少有提及，品牌方可抓住此痛点，与护肤虚拟社区协调在产品详情页加入购物平台官方旗舰店的链接与最新活动简介，拉动护肤虚拟社区用户向购物平台用户的转化。

参考文献：

[1] 王琳，梁孟华 . 基于知识发布的网络社区用户行为规律研究——以 Keep 健康社区为例 [J]. 情报探索，2022(1): 1-9.

[2] 杨瑞仙，张广轶 . 学术虚拟社区知识交流过程与机理研究 [J]. 现代情报，2020, 40(10): 52-61.

[3] 陈秀秀 . 基于微博评论的虚拟社区互动性对消费者品牌态度的影响及机制研究 [D]. 江门：五邑大学，2019.

[4] 谭春辉，李玥澎 . 基于用户评论与 Kano 模型的虚拟学术社区优化策略研究 [J]. 情报理论与实践，2021, 44(11): 108-115.

[5] 周阳，谭春辉，朱宸良，等 . 基于扎根理论的虚拟学术社区用户参与行为研究——以小木虫为例 [J]. 情报科学，2022, 40(1): 176-183.

[6] CHEN H, LIU F, WEN Y, et al. Compilation and application of the scale of sustainable knowledge sharing willingness in virtual academic community during the times of the coronavirus pandemic (COVID-19)[J]. Front psychol, 2021(12): 627833.

[7] TABISH M, BASHIR M A, ALAM M M, et al. The role of virtual community participation and engagement in building brand trust: evidence from Pakistan business schools[J]. The journal of Asian finance, economics and business, 2022, 9(3): 399-409.

[8] CARA P, BODKIN D. CHARLES. Online retail store community engagement and its impact on purchase intentions in the context of online reviews[J]. International journal of online marketing, 2022, 12(1): 1-15.

[9] CHAKRABORTY U. The impact of source credible online reviews on purchase intention[J]. Journal of research in interactive marketing, 2019, 13(2): 401-414.

[10] 屈慧君 . 互动活力与信任的链式中介效应——微博虚拟品牌社区用户参与动机对购买意愿的影响 [J]. 郑州大学学报 (哲学社会科学版), 2019, 52(2): 113-117.

[11] 郭恒睿，王中卿，朱巧明，等 . 基于半监督学习的中文社交文本事件聚类方法 [J]. 中文信息学报，2022, 36(2): 152-159.

[12] 谢红 . 基于词频比的改进 Jaccard 系数文本相似度计算 [J]. 内江科技，2021, 42(8): 27-28.

[13] 王春柳，杨永辉，邓霏，等 . 文本相似度计算方法研究综述 [J]. 情报科学，2019, 37(3): 158-168.

[14] 邱先标，陈笑蓉 . 一种基于特征加权的文本相似度计算算法 [J]. 贵州大学学报 (自然科学版), 2018, 35(1): 63-68.

[15] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation[J]. Journal of machine learning research, 2003, 3(4/5): 993-1022.

[16] HAJJEM M, LATIRI C. Combining IR and LDA topic modeling for filtering microblogs[J]. Procedia computer science, 2017(112): 761-770.

[17] 李璐萍，赵小兵 . 基于主题模型的主题发现方法研究综述 [J]. 中央民族大学学报 (自然科学版), 2021, 30(2): 59-66.

[18] 何琳，马晓雯，喻雪寒，等 . 典籍事件触发动词识别研究：基于《左传》的文本实验 [J]. 图书情报工作，2022, 66(5): 133-141.

[19] 王慧，孙国道，丁伟杰，等 . 司法判决文书数据中演化模式的挖掘和可视分析 [J]. 计算机辅助设计与图形学学报，2022, 34(3): 470-481.

[20] 张雷，谭慧雯，张璇，等 . 基于 LDA 模型的高校师德舆情演化及路径传导研究 [J]. 情报科学，2022, 40(3): 144-151.

[21] 崔宁，赵宗良，吴瑞雪 . 基于 LDA 主题模型和偏序集的在线商品评论研究 [J]. 情报探索，2021(12): 71-76..

[22] 黄仿元 . 利用机器学习算法实现对医院评价的情感分析 [J]. 信息技术与信息化，2021(2): 37-39.

[23] 尚永敏，赵榆琴 . 基于机器学习的在线评论情感分析与实现 [J]. 大理大学学报，2021, 6(12): 80-86.

[24] 艾媒数据中心 .2022 年中国洗面奶消费者行为洞察 [EB/OL].[2022-04-05]. https: //data.iimedia.cn/31026926/31026945/44290232.

[25] 21 经济网 . 微博回港二次上市：75% 用户为 90 后 社交“元老”开启第二春？ [EB/OL]. [2022-04-05]. http://www.21jingji.com/article/20211119/herald/55ae24eb120e9a397954803e832d4ada.html.

[26] CSDN. 折肘法 + 困惑度确定 LDA 主题模型的主题数 [EB/OL]. [2022-04-05]. https://blog.csdn.net/weixin_43343486/article/details/109255165.

[27] 池毛毛，潘美钰，王伟军 . 共享住宿与酒店用户评论文本的跨平台比较研究：基于 LDA 的主题社会网络和情感分析 [J]. 图书情报工作，2021, 65(2): 107-116.

知识管理论坛，2023(2): 79-91
DOI: 10.13266/j.issn.2095-5472.2023.007

A Comparative Study of Multi-Platform Consumer Review Topics Based on Co-Occurrence Network and Sentiment Analysis

Zhou Tingwei
School of Management, Shanghai University, Shanghai 200444

Abstract: [Purpose/Significance] The purpose of this article is to study the similarities and differences in the evaluation content of users of virtual life communities, social platforms and shopping platforms for the same product. **[Method/Process]** By selecting the three platforms of You Really Beautiful App, Weibo, and JD.com as the experimental objects, 54 071 user comment texts related to facial cleanser as basic skin care products were collected, using LDA topic generation model, co-occurrence network and machine learning-based sentiment analysis method, a multi-platform comparative analysis of user comment texts is carried out. **[Result/Conclusion]** The study found that there are similarities and differences in the topic feature words, co-occurrence network and topic sentiment of eight comment topics on the three platforms, and the content tendencies conform to the characteristics of each platform.

Keywords: multi-platform comparison text topic clustering co-occurrence network analysis emotion analysis